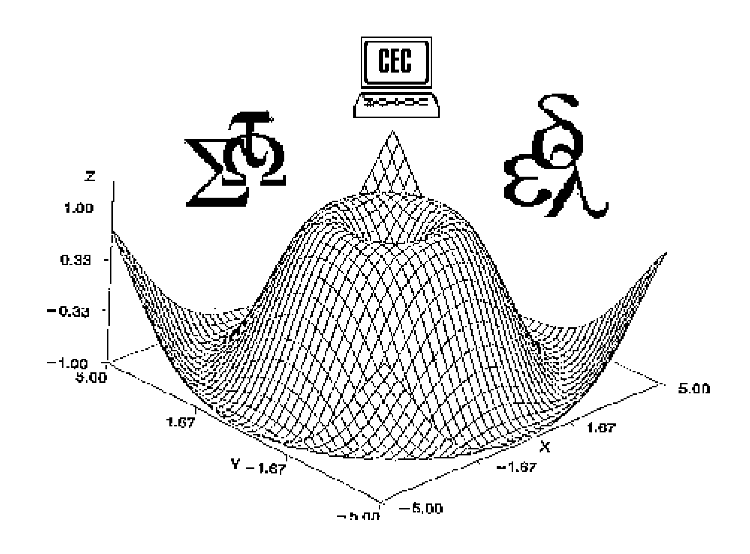
II FORO DE ESTADISTICA
I CONGRESO NACIONAL DE ESTADISTICA
I MEETING OF THE INTERNATIONAL BIOMETRIC SOCIETY, GUATEMALA GROUP



MINICURSO

ANALISIS DE EXPERIMENTOS CON EL SISTEMA SAS

M.Sc. VICTOR MANUEL ALVAREZ CAJAS
Ing. Agr. BYRON HUMBERTO GONZÁLEZ RAMÍREZ

ANALISIS DE EXPERIMENTOS CON EL SISTEMA SAS

M. Sc. Victor Manuel Alvarez Cajas
Profesor Titular, Centro de Estadística y Cálculo. Facultad de Agronomía

Ing. Agr. Byron Humberto González Ramírez
Administrador de Red, Centro de Estadística y Cálculo. Facultad de Agronomía

I. PRESENTACION

Partiendo del marco conceptual de lo que encierra el método científico podemos establecer el papel que juega la Estadística en el análisis y la interpretación de los resultados de experimentos.

El análisis y la interpretación objetiva de los resultados de un experimento puede considerarse tan importante como la conducción del mismo, y muchas veces esta parte es descuidada provocando una pérdida de recursos involucrados en la experimentación agrícola.

En atención a las premisas anteriores, las presentes notas pretenden dar lineamientos generales que permitan facilitar la elección de un diseño experimental que se ajuste más a las condiciones reales en que se requiere realizar el experimento, y además, sugerir la forma de cómo analizar tanto manual como electrónicamente los resultados que se obtengan.

El presente trabajo pretende cubrir los siguientes objetivos:

1. Que las presentes notas sean una ayuda complementaria a las personas que se relacionan con la experimentación agrícola.

2. Facilitar la elección de un diseño experimental en particular de acuerdo a las condiciones reales de campo.

3. Señalar la forma de ordenar los datos de campo y cómo realizar el análisis estadístico correspondiente.

4. Ilustrar la manera de aplicar métodos de análisis univariados a conjuntos de datos a través del Sistema -SAS-.

1. DISEÑO COMPLETAMENTE AL AZAR -DCA-

1.1 Características del Diseño.

La principal, es que los tratamientos se asignan a las unidades experimentales aleatoriamente (al azar) sin ninguna restricción. Su uso es muy frecuente cuando el experimento se llevará a cabo bajo condiciones homogéneas de humedad, temperatura, luz, suelo, fertilidad, entre otras.

1.2 Ventajas

a. Su análisis estadístico es fácil

b. Permite máxima flexibilidad en cuanto al número de tratamientos y número de repeticiones.

c. El análisis estadístico sigue siendo fácil aunque haya una pérdida de unidades experimentales.

d. El número de grados de libertad para el error es máximo en comparación con otros diseños.

1.3 Desventajas

a. Cuando el número de unidades experimentales es muy grande es difícil encontrar lugares grandes que presenten la homogeneidad requerida.

b. Generalmente otro diseño suele ser más eficiente (preciso).

1.4 Aleatorización

Por aleatorización se entiende la asignación al azar de los tratamientos a las unidades experimentales usando ya sea una tabla de números aleatorios, o una calculadora científica que posea la función RAN #.

Procedimiento

a. Determinar el número de unidades experimentales (n) y numerarlas. Es posible obtener ***n*** al multiplicar el número de tratamientos por el número de repeticiones. (n = t x r).

b. Asignar el número de unidades experimentales a cada tratamiento usando la tabla de números aleatorios o bien cualquier otro herramienta que sirva a este propósito. Ejemplo: Si cada tratamiento se repite cuatro veces; los 4 primeros números aleatorios obtenidos se asignan al tratamiento A; los siguientes cuatro números aleatorios al tratamiento B; y así sucesivamente.

c. Una vez hecha la distribución anterior (b), se numeran las unidades experimentales y se localizan los tratamientos de acuerdo al número que les corresponde y se obtiene así la distribución de campo. (croquis de campo).

1.5 Modelo Estadístico

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \qquad i = 1,2,\ldots t \quad j = 1,2,\ldots r$$

de donde,

Y_{ij}= Variable de respuesta de la ij-ésima unidad experimental

μ = Efecto de la media general

τ_i = Efecto del i - ésimo tratamiento

ε_{ij} = Efecto del error experimental asociado a la ij-ésima unidad experimental

1.6 Análisis de Varianza

1.6.1 Hipótesis

Ho: $\tau = \tau_i$ (Todos los tratamientos producen el mismo efecto)

Ha: $\tau \neq \tau_i$ para al menos un i; i = 1,2, . . . t. (Al menos uno de los tratamientos produce efectos distintos)

1.6.2 Tabla de Datos

Tratamientos	Repeticiones					$Y_{i.}$
	1	2	3	. . .	r	
1	Y_{11}	Y_{12}	Y_{13}	. . .	Y_{1r}	$Y_{1.}$
2	Y_{21}	Y_{22}	Y_{23}	. . .	Y_{2r}	$Y_{2.}$
3	Y_{31}	Y_{32}	Y_{33}	. . .	Y_{3r}	$Y_{3.}$
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
t	Y_{t1}	Y_{t2}	Y_{t3}	. . .	Y_{tr}	$Y_{t.}$

1.6.3 Supuestos:

Las suposiciones que validan el análisis de varianza son:

a. Los errores son independientes

b. Los errores están normalmente distribuidos con media cero y varianza constante

c. Existe homogeneidad de varianzas entre los tratamientos

1.6.4 Tabla de Análisis de Varianza

Fuentes de variación	Grados de libertad (gl)	Suma de Cuadrados	C. Medios	Fc
Tratamientos	t –1	$\sum_{i=1}^{r} \frac{Y_{i.}^2}{r} - \frac{Y_{..}^2}{tr}$	SC_{trat} / gl_{trat}	CM_{trat} / CM_{ee}
Error	t (r - 1)	SCtotal – Sctrat	SC_{ee} / gl_{ee}	
Total	t r - 1	$\sum_{i=1}^{t} \sum_{j=1}^{r} Y_{ij} - \frac{Y_{..}^2}{tr}$		

Regla de Decisión

Rechazar Ho. si $Fc \geq Ft_{(gl\ trat;\ gl\ error;\ \alpha)}$
No Rechazar Ho. si $Fc < Ft_{(gl\ trat;\ gl\ error;\ \alpha)}$

Ft = Valor tabular que puede ser encontrado en la tabla 1 del anexo, considerando gl trat y gl error.

1.6.5 Coeficiente de Variación (CV)

Se le puede considerar como medida relativa de la variación que no es posible controlar en el experimento (error experimental) y se calcula de la siguiente forma:

$$CV = \frac{\sqrt{CM_{ee}}}{\bar{Y}_{..}} * 100 \qquad \bar{Y}_{..} = \frac{Y_{..}}{tr}$$

1.7 Ejemplo de Aplicación

Se probaron concentraciones de 1, 2, 3 y 4 ppm de una cierta enzima mas un control (en que no se añadió enzima) para estudiar su efecto en la separación de un jugo de naranja. Cada tratamiento se replicó 4 veces y se usó un diseño completamente al azar. Se añadió agua a las 20 muestras y se midió el tiempo de separación (minutos) en cada muestra. Los resultados pueden observarse en el cuadro 1. Interesa responder a las siguientes preguntas:

1. La presencia de enzima retarda la separación comparada con la ausencia de la misma?
2. Hay algún efecto diferencial del nivel de enzima agregado?

Cuadro 1: Resultados en minutos, del tiempo de separación de un jugo de naranja luego de la aplicación de una enzima.

Tratamiento	Minutos
0	3.96
0	6.24
0	5.42
0	11.11
1	27.31
1	26.96
1	32.21
1	30.13
2	35.3
2	34.71
2	36.75
2	38.38
3	41.09
3	43.99
3	48.7
3	41.78
4	48.9
4	47.61
4	50.38
4	49.61

1.7.1 Programa de SAS

A continuación el programa SAS y la salida que responde a las preguntas anteriores y además prueba la normalidad de los residuales.

```
OPTIONS NODATE;
DATA dos;
INPUT trat minutos;
CARDS;
0 3.96
0 6.24
.
.
4 47.61
4 50.38
4 49.61
;
PROC GLM;
    CLASS trat;
    MODEL minutos=trat;
    MEANS trat/TUKEY;
    OUTPUT OUT=res r=rmin;
RUN;
PROC PRINT DATA=res;
RUN;
PROC UNIVARIATE PLOT NORMAL DATA=res;
    VAR rmin;
RUN;
```

1.7.2 Salida de SAS

```
                                SAS

                    General Linear Models Procedure
                       Class Level Information

                      Class     Levels     Values

                      TRAT           5     0 1 2 3 4


                Number of observations in data set = 20

                                SAS

                    General Linear Models Procedure

Dependent Variable: MINUTOS
                                      Sum of          Mean
Source                    DF         Squares        Square   F Value     Pr > F

Model                      4     4387.237470   1096.809367    173.25     0.0001

Error                     15       94.960550      6.330703

Corrected Total           19     4482.198020

               R-Square             C.V.       Root MSE         MINUTOS Mean

               0.978814         7.618279       2.516089           33.0270000



                               SAS
```

```
                      General Linear Models Procedure

Dependent Variable: MINUTOS

Source                  DF       Type I SS    Mean Square   F Value     Pr > F

TRAT                     4     4387.237470    1096.809368    173.25     0.0001

Source                  DF     Type III SS    Mean Square   F Value     Pr > F

TRAT                     4     4387.237470    1096.809368    173.25     0.0001
                                    SAS

                      General Linear Models Procedure

          Tukey's Studentized Range (HSD) Test for variable: MINUTOS
         NOTE: This test controls the type I experimentwise error rate, but
               generally has a higher type II error rate than REGWQ.

                     Alpha= 0.05  df= 15  MSE= 6.330703
                 Critical Value of Studentized Range= 4.367
                   Minimum Significant Difference= 5.4938
         Means with the same letter are not significantly different.

                Tukey Grouping              Mean      N  TRAT

                             A            49.125      4  4
                             A            43.890      4  3
                             B            36.285      4  2
                             C            29.152      4  1
                             D             6.682      4  0


                                  SAS

                     OBS    TRAT    MINUTOS      RMIN

                       1      0       3.96     -2.7225
                       2      0       6.24     -0.4425
                       3      0       5.42     -1.2625
                       4      0      11.11      4.4275
                       5      1      27.31     -1.8425
                       6      1      26.96     -2.1925
                       7      1      32.21      3.0575
                       8      1      30.13      0.9775
                       9      2      35.30     -0.9850
                      10      2      34.71     -1.5750
                      11      2      36.75      0.4650
                      12      2      38.38      2.0950
                      13      3      41.09     -2.8000
                      14      3      43.99      0.1000
                      15      3      48.70      4.8100
                      16      3      41.78     -2.1100
                      17      4      48.90     -0.2250
                      18      4      47.61     -1.515
                      19      4      50.38      1.255
                      20      4      49.61      0.485


                                   SAS
                           UNIVARIATE PROCEDURE

Variable=RMIN
                                  Moments


                   N                20  Sum Wgts         20
                   Mean              0  Sum               0
                   Std Dev    2.235604  Variance   4.997924
                   Skewness   0.838511  Kurtosis   -0.02402
                   USS        94.96055  CSS        94.96055
                   CV                .  Std Mean   0.499896
                   T:Mean=0          0  Prob>|T|     1.0000
                   Sgn Rank        -11  Prob>|S|     0.7012
                   Num ^= 0         20
                   W:Normal   0.921031  Prob<W       0.1096



                                   SAS
                           UNIVARIATE PROCEDURE
Variable=RMIN
                              Quantiles(Def=5)
                    100% Max      4.81        99%      4.81
                     75% Q3    1.11625        95%   4.61875
                     50% Med  -0.33375        90%    3.7425
```

```
                      25% Q1    -1.70875        10%    -2.4575
                       0% Min       -2.8         5%  -2.76125
                                                 1%      -2.8
                     Range          7.61
                     Q3-Q1         2.825
                     Mode           -2.8

                                    SAS
                             UNIVARIATE PROCEDURE
Variable=RMIN
                                  Extremes
                       Lowest    Obs      Highest    Obs
                          -2.8(       13)     1.255(       19)
                      -2.7225(        1)     2.095(       12)
                      -2.1925(        6)    3.0575(        7)
                         -2.11(       16)    4.4275(        4)
                      -1.8425(        5)      4.81(       15)


                                    SAS
                             UNIVARIATE PROCEDURE
Variable=RMIN
                Stem Leaf                     #             Boxplot
                   4 48                       2                |
                   2 11                       2                |
                   0 15503                    5             +--+--+
                  -0 8653042                  7             *-----*
                  -2 8721                     4                |
                     ----+----+----+----+

                                    SAS

                             UNIVARIATE PROCEDURE

Variable=RMIN

                                Normal Probability Plot
                 5+                                        *    +*++++++
                  |                                    *+*++++++
                 1+                          ++***+**+
                  |                     +**+**** *
                -3+       * +++*++*+*
                   +----+----+----+----+----+----+----+----+----+----+
                       -2        -1         0        +1        +2
```

2. DISEÑO DE BLOQUES COMPLETOS AL AZAR

2.1 Características del diseño

Es el más usado en la experimentación agrícola, y es aquel en que:

a. Es conveniente su empleo cuando se logra determinar una gradiente de variabilidad en un sentido, pudiendo ser esta climática, edáfica, o de otro tipo.

b. Las unidades experimentales se distribuyen en grupos o bloques, de tal manera que las unidades experimentales dentro de un bloque sean relativamente homogéneas, y además el número de unidades experimentales dentro de un bloque debe ser igual al número de tratamientos a investigar.

c. Los tratamientos se asignan al azar a las unidades experimentales dentro de cada bloque, de manera independiente

2.2 Algunas recomendaciones para su uso

i. Dividir el lugar donde se llevará a cabo el experimento en grupos homogéneos (bloques). Cada bloque completo será una repetición.

ii. Los bloques se colocarán en forma perpendicular a la gradiente que se ha establecido.

iii. Dividir el bloque en tantas unidades experimentales como tratamientos se quieran estudiar. Cada tratamiento debe aparecer una sola vez en cada bloque.

iv. Tratar que la parcela experimental quede localizada de tal manera que su longitud más grande esté orientada paralelamente a la gradiente.

v. Sortear independientemente en cada bloque los distintos tratamientos.

vi. En el manejo del experimento, tratar que éste sea lo más homogéneo posible. En otras palabras, realizar en un mismo día actividades como fertilización, riego, aplicación de fungicidas, etc.; lo ideal sería para el experimento completo; de no poderse, efectuar las actividades por bloques completos. En conclusión no hacer trabajos que dejen bloques incompletos.

2.3 Ventajas

a. Al usar un bloques al azar se obtienen resultados más precisos que cuando se usa un completamente al azar.

b. Puede incluirse cualquier número de tratamientos y repeticiones hasta donde sea prácticamente posible. (Es recomendable que el número de tratamientos evaluados no sea superior a 15)

c. Análisis estadístico fácil.

d. Si se pierden unidades experimentales (hasta 3) el análisis estadístico sigue siendo fácil. (Ver datos perdidos sección 2.8)

e. Si se pierde un tratamiento completo se analiza como un tratamiento menos; si es un bloque completo, como una repetición menos.

2.4 Desventajas

a. Hay menor número de grados de libertad para el error experimental respecto al diseño completamente al azar.

2.5 Aleatorización

Cuando las unidades experimentales han sido agrupadas en bloques; los tratamientos se asignan aleatoriamente a las unidades dentro de cada bloque.

Para hacer ésta asignación se puede usar el Procedimiento *PLAN* de SAS (PROC PLAN), un simple sorteo, o bien usando una tabla de números aleatorios. Ejemplo: si fueran cuatro los tratamientos a evaluar, numerarlos correlativamente, y luego con la tabla de números aleatorios ubicar en la primera unidad experimental el primer número que aparezca y que esté comprendido entre 1 y 4, hasta completar los cuatro números. El mismo procedimiento para los siguientes bloques.

2.6 Modelo Estadístico

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \qquad i = 1,2, \ldots t \quad j = 1,2, \ldots r$$

de donde,

Y_{ij} = Variable de respuesta de la ij-ésima unidad experimental

μ = Efecto de la media general

τ_i = Efecto del i - ésimo tratamiento

β_j= Efecto del j - ésimo bloque

ε_{ij}= Efecto del error experimental asociado a la ij-ésima unidad experimental

2.7. Análisis de Varianza

2.7.1 Hipótesis

Ho: $\tau = \tau_i$ (Todos los tratamientos producen el mismo efecto)

Ha: $\tau \neq \tau_i$ para al menos un i; i = 1,2, . . . t. (Al menos uno de los tratamientos produce Efectos distintos)

2.7.2 Tabla de Datos

Tratamientos	Repeticiones					
	1	2	3	. . .	r	
1	Y_{11}	Y_{12}	Y_{13}	. . .	Y_{1r}	$Y_{1.}$
2	Y_{21}	Y_{22}	Y_{23}	. . .	Y_{2r}	$Y_{2.}$
3	Y_{31}	Y_{32}	Y_{33}	. . .	Y_{3r}	$Y_{3.}$
.	.	.	.	. . .	.	.
.	.	.	.	. . .	.	.
t	Y_{t1}	Y_{t2}	Y_{t3}	. . .	Y_{tr}	$Y_{t.}$
$Y_{.j}$	$Y_{.1}$	$Y_{.2}$	$Y_{.3}$	. . .	$Y_{.r}$	$Y_{..}$

2.7.3 Tabla de Análisis de Varianza

FV	GL	SC	CM	Fc
Bloques	r- 1	$\sum_{j=1}^{r} \frac{Y_{.j}^2}{t} - \frac{Y_{..}^2}{tr}$		
Tratamientos	t – 1	$\sum_{i=1}^{t} \frac{Y_{i.}^2}{r} - \frac{Y_{..}^2}{tr}$	SC_{trat} / gl_{trat}	CM_{trat}/CM_{ee}
Error exp.	(t -1) (r -1)	SCtotal-SCtrat - SC bloque	SCee /gl ee	
Total	tr - 1	$\sum_{i=1}^{t} \sum_{j=1}^{r} Y_{ij} - \frac{Y_{..}^2}{tr}$		

2.7.4 Regla de Decisión

Previo a decidir sobre el rechazo o no de la hipótesis planteada deberá desarrollarse el análisis de varianza completo, cuya tabla guía se presenta en 2.7.4.

Rechazar Ho. si $Fc \geq Ft_{(gl\ trat;\ gl\ error;\ \alpha)}$
No Rechazar Ho. si $Fc < Ft_{(gl\ trat;\ gl\ error;\ \alpha)}$

Ft = Valor tabular que puede ser encontrado en la tabla 1 del apéndice, considerando gl trat y gl error.

2.8 Datos Perdidos

Frecuentemente en la práctica por razones ajenas puede ocurrir:

a. Que se pierda un bloque completo
b. Que se pierda un tratamiento completo
c. Que se pierdan unidades experimentales.

Qué hacer en cado uno de estos casos:

a. Si el perdido es un bloque completo realizar el análisis con una repetición menos.

b. Si se pierde un tratamiento completo analizar el experimento con un tratamiento menos.

c. Si existen unidades experimentales perdidas, es posible hacer una estimación de sus valores mediante los siguientes procedimientos:

2.8.1 Una (1) Unidad Perdida.

El procedimiento en éste caso es sencillo ya que basta aplicar las siguiente fórmula para estimar el valor perdido. (Ver nota importante pagina 15).

de donde:

Yij = dato estimado o perdido

r = número de bloques

$$Y_{ij} = \frac{rB + tT - S}{(r-1)\ (t-1)}$$

B = Suma de las observaciones en el mismo bloque de la observación perdida.

t = Número de tratamientos.

T = Suma de las observaciones con el mismo tratamiento de la observación perdida.

$$a = \frac{\bar{Y}_{i.} + \bar{Y}_{.j}}{2} \qquad b = \frac{\bar{Y}_{i.} + \bar{Y}_{.j}}{2}$$

S = Suma de todas las observaciones reales.

2.8.2 Dos (2) Unidades Perdidas:

Para estimar el valor para dos estimaciones perdidas se emplea el *método iterativo* que consiste en la aplicación de las siguientes fórmulas:

= Promedio del tratamiento con la unidad perdida
= Promedio del bloque con la unidad perdida

$$a = \frac{\bar{Y}_{i.} + \bar{Y}_{.j}}{2} \qquad b = \frac{rB + tT - S}{(r-1)\ (t-1)}$$

$\bar{Y}_{i.}$
$\bar{Y}_{.j}$

Teniendo los datos ordenados se identifican arbitrariamente las dos unidades perdidas como a ó b, y luego se aplica la fórmula para "a" y después usamos la fórmula para "b"; enseguida omitimos el valor estimado en "a" y volvemos a aplicar "b", repetimos el mismo proceso omitiendo el nuevo valor de "b". Detenemos este método iterativo hasta que los valores se repitan y los dos últimos serán los valores a usar para el posterior análisis.(Ver nota importante página 15).

2.8.3 Tres (3) Unidades Perdidas

Se emplea también el método iterativo; usando las siguientes fórmulas:

Con la tabla de datos ya ordenada identificar arbitrariamente con las letras *a, b* y *c* las unidades experimentales perdidas y aplicarles a cada una las fórmulas correspondientes. Hecho lo anterior omitir el valor encontrado de "a" y volverlo a calcular usando "c". Hacer los mismo para "b" y "c" originales; repetir el ciclo las veces que sean necesarias hasta que los valores para cada uno se repitan. Encontrados los valores proseguir con el análisis. (Ver la siguiente nota

$$c = \frac{rB + tT - S}{(r-1)\ \ (t-1)}$$

importante).

Nota Importante:

a. Lo descrito anteriormente tiene justificación siempre y cuando las unidades experimentales estén localizadas en bloques y tratamientos diferentes.

b. Al realizar el análisis de varianza; tener presente que debemos restarle a los grados de libertad del total y del error experimental un número que es igual al número de unidades experimentales estimadas *(k)* así:

gl Total = tr - 1 - k
gl error = (t - 1) (r - 1) - k

c. Al realizar la prueba de Tukey o SNK usar el comparador W' cuando se esté comparando el tratamiento al cual se le estimó el valor, con cualquier otro:

q = valor tabular que aparece en la tabla de Tukey
p = número de tratamientos

$$W = q_{(p,glee,a)} * S\bar{x} \qquad S\bar{x} = \sqrt{CME\left[\frac{2}{r} + \frac{t}{r(r-1)\,(t-1)}\right]}$$

2.9 Ejemplo de Aplicación

Con el propósito de comparar el efecto de cinco fuentes diferentes de nitrógeno y un control sobre la producción de materia seca de cebada forrajera, se condujo un experimento empleando un diseño en bloques completos al azar. Las cinco fuentes evaluadas fueron:

1. $(NH_4)_2$ 2. $NH_4\ NO_3$ 3. $CO\ (NH_4)_3$
4. $Ca\ (NO_3)_2$ 5. $Na\ NO_3$ 6. Control (sin fertilizante)

Dado que se deseaba aplicar los resultados sobre condiciones relativamente amplias, se decidió realizar el experimento en cuatro tipos de suelo. De esta manera, el diseño experimental usado fue el de bloques al azar para asi poder controlar con los bloques el efecto de los diferentes tipos de suelo. A la madurez del cultivo, se cosecharon las diferentes unidades experimentales y se midió la cantidad de materia seca de forraje producido. Las producciones (en kilogramos por parcela) se presentan en el cuadro 2.

Cuadro 2: Producción (kilogramos por parcela) de cebada de los seis tratamientos nitrogenados.

Tratamientos	Bloques			
	I	II	III	IV
1	32.1	35.6	41.9	35.4
2	30.1	31.5	37.1	30.8
3	25.4	27.4	33.8	31.1
4	24.1	33.0	35.6	31.4
5	26.1	31.0	33.8	31.9
6	23.2	24.8	26.7	26.7

2.9.1 Programa de SAS

A continuación el programa SAS y la salida que considera el análisis de varianza de dos vías.

```
OPTIONS NODATE;
DATA dos;
INPUT bloque $ fuente matsec;
CARDS;
I   1   32.1
I   2   30.1
I   3   25.4
I   4   24.1
```

```
.
.
IV  4  31.4
IV  5  31.9
IV  6  26.7
;
PROC ANOVA;
    CLASS fuente bloque;
    MODEL matsec= bloque fuente;
    MEANS fuente/TUKEY;
RUN;
```

2.9.2 Salida de SAS

```
                              The SAS System

                      Analysis of Variance Procedure
                         Class Level Information

                      Class    Levels    Values

                      FUENTE        6    1 2 3 4 5 6

                      BLOQUE        4    I II III IV


                 Number of observations in data set = 24

                              The SAS System

                      Analysis of Variance Procedure

Dependent Variable: MATSEC
                                     Sum of            Mean
Source                  DF          Squares          Square   F Value     Pr > F

Model                    8     447.83166667     55.97895833     18.86     0.0001

Error                   15      44.52791667      2.96852778

Corrected Total         23     492.35958333

                  R-Square             C.V.        Root MSE          MATSEC Mean

                  0.909562         5.584146       1.7229416            30.854167



Source                  DF         Anova SS     Mean Square   F Value     Pr > F

FUENTE                   5     255.27708333     51.05541667     17.20     0.0001
BLOQUE                   3     192.55458333     64.18486111     21.62     0.0001




                              The SAS System

                      Analysis of Variance Procedure

          Tukey's Studentized Range (HSD) Test for variable: MATSEC

          NOTE: This test controls the type I experimentwise error rate, but
                generally has a higher type II error rate than REGWQ.

                     Alpha= 0.05  df= 15  MSE= 2.968528
                  Critical Value of Studentized Range= 4.595
                    Minimum Significant Difference= 3.9582

          Means with the same letter are not significantly different.
```

Tukey Grouping		Mean	N	FUENTE
	A	36.250	4	1
B	A	32.375	4	2
B		31.025	4	4
B		30.700	4	5
B		29.425	4	3
	C	25.350	4	6

3. COMPARACION MULTIPLE DE MEDIAS

El análisis de varianza indica si existe o no variación entre los tratamientos y una vez encontrada ésta variación el procedimiento que señala qué tratamiento (s) está (n) causando esa variación es la comparación múltiple de medias; cuyas pruebas más usuales son: Tukey, Duncan, SNK; aunque también pueden ser usadas pruebas como Bonferroni, Scheffé, Deponly, Bon, Dunnett, Gabriel, Sidak, Waller, u otras según sean las necesidades del problema de aplicación.

3.1 Prueba de Tukey

Procedimiento

a. Calcular el comparador W.

de donde:

q = valor que aparece en la tabla de Tukey

glee = grados de libertad del error

$$W = q_{(p, glee, a)} * S\bar{x}$$

p = número de tratamientos

$$S\bar{x} = \sqrt{\frac{CME}{r}}$$

Sx = Error estándar de la media

r = número de repeticiones

$$Wp = q_{(p,glee,a)} * S\bar{x} \qquad S\bar{x} = \sqrt{\frac{CME}{r}}$$

α = nivel de significancia.

b. *Encontrar las diferencias entre las medias*

Para encontrar todas las posibles diferencias es necesario hacer un cuadro de doble entrada colocando horizontalmente las medias ordenadas de mayor a menor y verticalmente de menor a mayor; llenando el cuadro con las respectivas diferencias. (d_i)

c. Regla de Decisión

Encontradas las diferencias de las medias, comparar cada una de ellas con W, tomando en consideración lo siguiente:

Hay diferencia significativa si $|d_i| \geq W$ (*)
No hay diferencia significativa si $|d_i| < W$ (ns)

$|d_i|$ = valor absoluto de la diferencia entre dos tratamientos.

d. Presentación de medias

Esta presentación se hace ordenando verticalmente las medias de mayor a menor. Unir luego con una misma letra o línea, aquellas medias que son significativamente diferentes (ns).

3.2 Prueba de SNK

Es una modificación de la prueba de Tukey; en donde se usan varios comparadores.

Procedimiento:

a. Cálculo de comparadores (Wp)

donde:

q = valor que aparece en la tabla de Tukey

glee = grados de libertad para el error

i = 1,2,3, . . . t

r = número de repeticiones

b. *Encontrar las diferencias entre las medias:* igual a Tukey.

c. *Regla de Decisión*

Esta prueba no exige que la diferencia entre dos medias sea tan grande como sucede en Tukey para declarar significativa la diferencia. A mayor diferencia entre los tratamientos corresponde un mayor comparador, manteniendo esta regla de decisión.

Hay diferencia significativa si $|d_i| \geq W$ (*)
No hay diferencia significativa si $|d_i| < W$ (ns)

$|d_i|$ = valor absoluto de la diferencia entre dos tratamientos.

d. *Presentación de Medias*

Igual a Tukey.

4. DISEÑO CUADRADO LATINO

4.1 Características

En este diseño se impone a las unidades experimentales una restricción del tipo de doble bloqueo (hileras y columnas), es decir, la agrupación es en dos sentidos. Los tratamientos se agrupan en repeticiones de dos maneras distintas; cada hilera y cada columna constituye una repetición completa.

Por lo tanto, cada tratamiento debe aparecer una y solo una vez por columna y por hilera de tal manera que cualquier comparación de tratamientos no se vea afectada por las diferencias existentes entre hileras o entre columnas. En este diseño el número de hileras (h), número de columnas (c) y el número de tratamientos (t) debe ser igual. De tal manera que el número total de unidades experimentales será igual a:

$$h^2 \text{ ó } c^2 \text{ ó } t^2$$

Por lo anterior el uso de este diseño se recomienda para conducir experimentos en condiciones heterogéneas siempre y cuando podamos bloquear en dos sentidos. Ejemplo en el campo:

fertilidad - salinidad
fertilidad - sembradora
jornada de trabajo - edad del operario

4.2 Ventajas

a. Análisis estadístico fácil

b. Controla más las fuentes de variación que el diseño en bloques al azar, pues controla la variación de hileras, columnas y tratamientos; o sea que disminuye el error experimental al extraer estas tres fuentes de variación.

Para aprovechar las ventajas de ésta distribución, es indispensable lo siguiente:

a. Dividir el lote, o lugar de la experiencia en un número de unidades experimentales igual al cuadrado del número de tratamientos.
b. El número de repeticiones debe ser igual al número de tratamientos.

c. Formar hileras y columnas de unidades experimentales iguales al número de repeticiones y de tratamientos.

d. Distribuir los tratamientos en tal forma que ningún tratamiento se repita en fila ni en columna (ver aleatorización).

4.3 Desventajas

a. Disminuye el número de grados de libertad para el error experimental.

b. Sólo se puede usar cuando se tiene un número limitado de tratamientos (usualmente entre 4 y 8).

4.4. Aleatorización

Cuadrado Latino Reducido:

Es aquel en el cual la primera hilera y la primera columna están dispuestas en orden alfabético.

Procedimiento

a. Seleccionar un cuadrado latino reducido

b. Permutar hileras aleatoriamente

c. Permutar columnas aleatoriamente

d. Asignar los tratamientos a las letras al azar.

Los incisos *b,c,d* se pueden realizar usando una tabla de números aleatorios, o bien a través de un simple sorteo.

4.5 Modelo Estadístico

$$Y_{ijk} = \mu + H_i + C_j + \tau_k + \varepsilon_{ijk}$$

$i = 1,2,3, \ldots h = n$
$j = 1,2,3, \ldots c = n$ $\qquad h = c = t = n$
$k = 1,2,3, \ldots t = n$

de donde:

Y_{ijk} = Variable de respuesta de la ijk - ésima unidad experimental
μ = Efecto de la media general
H_i = Efecto de la i - ésima fila
C_j = Efecto de la j - ésima columna
τ_k = Efecto del k - ésimo tratamiento
ε_{ijk}= Error experimental asociado a la ijk - ésima unidad experimental

4.6 Análisis de Varianza

4.6.1 Hipótesis

Ho: $\tau = \tau_i$ (Todos los tratamientos producen el mismo efecto)
Ha: $\tau \neq \tau_i$ para al menos un i; i = 1,2, . . . t. (Al menos uno de los tratamientos produce efectos distintos)

4.6.2 Tabla de Datos *(para un cuadrado latino 4*4)*

Filas	Columnas				
	1	2	3	4	Total hileras $Y_{i..}$
1	Y_{11k}	Y_{12k}	Y_{13k}	Y_{14k}	$Y_{1....}$
2	Y_{12k}	Y_{22k}	Y_{23k}	Y_{24k}	$Y_{2..}$
3	Y_{13k}	Y_{32k}	Y_{33k}	Y_{34k}	$Y_{3..}$
4	Y_{14k}	Y_{42k}	Y_{43k}	Y_{44k}	$Y_{4..}$
Total columnas $Y_{.j.}$	$Y_{.1..}$	$Y_{.2..}$	$Y_{.3..}$	$Y_{.4..}$	$Y_{.....}$ (Gran total)

Donde k=1,2,3,4.

La tabla anterior se obtiene sin modificar el croquis de campo. Puede generarse una tabla adicional para los tratamientos, sumando todas las veces que se repitió cada tratamiento, obteniendo asi el respectivo total ($Y_{..k}$), como se indica a continuación:

Tratamientos			
1	2	3	4
Y_{ij1}	Y_{ij2}	Y_{ij3}	Y_{ij4}
Y_{ij1}	Y_{ij2}	Y_{ij3}	Y_{ij4}
Y_{ij1}	Y_{ij2}	Y_{ij3}	Y_{ij4}
Y_{ij1}	Y_{ij2}	Y_{ij3}	Y_{ij4}

$$CV = \frac{\sqrt{CME}}{\bar{Y}_{...}} * 100$$

Total Tratamientos $Y_{..k}$	$Y_{..1}$	$Y_{..2}$	$Y_{..3}$	$Y_{..4}$
Media Tratamientos $\bar{Y}_{..k}$	$\bar{Y}_{..1}$	$\bar{Y}_{..2}$	$\bar{Y}_{..3}$	$\bar{Y}_{..4}$

Media general = $\bar{Y}_{...} = \frac{Y_{...}}{t^2}$

4.6.3 Tabla de Análisis de Varianza

FV	gl	SC	CM	Fc
Tratamientos	t - 1	$\frac{\sum_{k=1}^{t} Y_{..k}^2}{t} - \frac{Y_{...}^2}{t^2}$	SC_{trat}/gl_{trat}	CM_{trat}/CM_{ee}
Hileras	t - 1	$\frac{\sum_{k=1}^{t} Y_{..i}^2}{t} - \frac{Y_{...}^2}{t^2}$		
Columnas	t - 1	$\frac{\sum_{k=1}^{t} Y_{.j.}^2}{t} - \frac{Y_{...}^2}{t^2}$		
Error exp.(ee)	(t -1) (t - 2)	SC total- {SC trat + SCH + SCC}	SCee /gl ee	
Total	$t^2 - 1$	$\sum_{i=1}^{t} \sum_{j=1}^{t} \sum_{k=1}^{t} Y_{ijk}^2 - \frac{Y_{...}^2}{t^2}$		

4.6.4 Regla de Decisión

Rechazar Ho. si $Fc \geq Ft_{(gl\ trat;\ gl\ error;\ \alpha)}$

No Rechazar Ho. si $Fc < Ft_{(gl\ trat;\ gl\ error;\ \alpha)}$

4.7 Datos Faltantes.

Cuando se tiene una unidad experimental perdida se usa la fórmula que se da a continuación obtenida por un procedimiento similar al de bloques al azar, es decir minimizando la suma de cuadrados del error.

$$a = \frac{\bar{Y}_{i..} + \bar{Y}_{.j.} + \bar{Y}_{..k}}{3}$$

$$Y_{ijk} = \frac{t\,(H + C + T) - 2S}{(t-1)\,(t-2)}$$

Donde;

t = número de tratamientos

H = total de valores observados para la hilera que contiene la unidad faltante

C = total de valores observados para la columna que contiene la unidad faltante

T = total del tratamiento que contiene la unidad faltante

S = gran total de valores reales observados.

Luego se efectúa el ANDEVA, restando uno (1) a los grados de libertad del total y del error experimental. Cuando se efectúan las comparaciones de medias de tratamientos se deben observar las mismas precauciones mencionadas para el diseño en bloques al azar ; es decir, que cuando se esté comparando el tratamiento al cual le estimamos el valor se deberá usar el error estandar siguiente:

$$S\bar{x} = \sqrt{CME\left[\frac{2}{t} + \frac{t}{(t-1)\,(t-2)}\right]}$$

Si los valores perdidos son dos o más usar el método iterativo (descrito en bloques al azar) empleando las siguientes fórmulas:

b. La fórmula para una unidad experimental perdida

4.8 Ejemplo de aplicación

Un ensayo con el propósito de evaluar el efecto de cinco productos químicos sobre la longitud (cm) de la parte aérea de plántulas de café noventa días después de la siembra, incluyó los siguientes productos:

A: Nitrato de potasio
B: Thiourea
C: Mezcla 1: Biozime T.S.
D: Mezcla 2: Bellgro PGR IV
E: Testigo (sin aplicación)

Los tratamientos se distribuyeron mediante un diseño en cuadrado latino, y los resultados de la evaluación se aprecian en el cuadro 3.

Cuadro 3: Valores de longitud (cm) de la parte aérea de plánulas de café.

E	A	C	D	B
2.4	2.4	1.9	2.4	2.0
B	C	E	A	D
1.7	2.4	2.7	3.6	3.0
A	B	D	E	C
2.7	2.1	3.8	2.6	1.8
D	E	B	C	A
2.6	3.1	2.3	2.2	2.6
C	D	A	B	E
2.9	3.1	3.0	2.0	2.2

4.8.1 Programa de SAS

A continuación el programa SAS y la salida que considera el análisis de varianza en cuadrado latino.

```
OPTIONS NODATE;
DATA dos;
INPUT columna fila trat $ long;
CARDS;
1 1 A 2.4
1 2 B 1.7
1 3 C 2.7
1 4 D 2.6
1 5 E 2.9
2 1 B 2.4
2 2 C 2.4
```

```
.
.
5 2 A 3.0
5 3 B 1.8
5 4 C 2.6
5 5 D 2.2
;
PROC GLM;
    CLASS columna fila trat;
    MODEL long= columna fila trat;
    MEANS trat/TUKEY;
RUN;
```

4.8.2 Salida de SAS

```
                              The SAS System

                      Analysis of Variance Procedure
                         Class Level Information

                       Class    Levels    Values

                       COLUMNA       5    1 2 3 4 5

                       FILA          5    1 2 3 4 5

                       TRAT          5    A B C D E

                Number of observations in data set = 25

                              The SAS System

                      Analysis of Variance Procedure

Dependent Variable: LONG
                                    Sum of            Mean
Source                  DF         Squares          Square   F Value     Pr > F

Model                   12      4.48800000      0.37400000      1.98     0.1263

Error                   12      2.27200000      0.18933333

Corrected Total         24      6.76000000


                  R-Square            C.V.        Root MSE            LONG Mean

                  0.663905        17.13089       0.4351245            2.5400000


Source                  DF        Anova SS     Mean Square   F Value     Pr > F

COLUMNA                  4      0.50800000      0.12700000      0.67     0.6247
FILA                     4      0.68000000      0.17000000      0.90     0.4951
TRAT                     4      3.30000000      0.82500000      4.36     0.0209

                             The SAS System
                      Analysis of Variance Procedure

            Tukey's Studentized Range (HSD) Test for variable: LONG

          NOTE: This test controls the type I experimentwise error rate, but
                generally has a higher type II error rate than REGWQ.

                       Alpha= 0.05  df= 12  MSE= 0.189333
                   Critical Value of Studentized Range= 4.508
                     Minimum Significant Difference= 0.8772

          Means with the same letter are not significantly different.
```

Tukey Grouping		Mean	N	TRAT
	A	2.9800	5	D
B	A	2.8600	5	A
B	A	2.6000	5	E
B	A	2.2400	5	C
B		2.0200	5	B

5. EXPERIMENTOS FACTORIALES

5.1 Conceptos

5.1.1 Experimentos Factoriales

Son aquellos en donde se estudian simultáneamente dos o más factores; cada uno incluyendo una serie de niveles.

Los experimentos factoriales no son diseños de campo sino simplemente una modalidad de estudio simultáneo de tratamientos; la combinación o arreglo de tratamientos se puede distribuir en bloques al azar, completamente al azar o cuadrado latino.

El arreglo combinatorio de tratamientos consiste en hacer todas las combinaciones posibles entre los niveles de cada uno de los factores estudiados y considerar cada combinación como un tratamiento.

Factor:

Un factor es un conjunto de tratamientos que se pueden aplicar a las unidades experimentales. Generalmente se denotan con letras mayúsculas (A,B, . . . Z o con las iniciales de los factores a probar). Los factores pueden ser cuantitativos (cantidad de fertilizante, de insecticida, una hormona, de tiempo, temperatura, concentración, etc.) o cualitativos (variedades, métodos de aplicación, marcas de producto, razas, procedencia, etc.).

Nivel:

Un nivel de un factor es un tratamiento particular de un conjunto de tratamientos que constituyen un factor. Si el factor es cuantitativo, los niveles están formados por las cantidades o dosis del mismo; si el factor es cualitativo los niveles los constituyen las manifestaciones del mismo o los tratamientos dentro del factor (nombres de las variedades, diferentes métodos de aplicación,las diferentes marcas del producto, las diferentes razas, etc.). Por lo general se identifican con letras minúsculas (a_1, a_2, . . . ; b_1, b_2, . . .; v_1, v_2, . . .; etc.)

Efecto simple:

El efecto de un factor a un solo nivel de otro factor o en una sola combinación de otros factores se conoce como ***efecto simple***. Si se promedian esos efectos simples se obtiene un efecto principal, y de manera más general, el efecto principal de un factor es una comparación entre respuestas esperadas para los diferentes niveles de un factor, promediados sobre todos los niveles de todos los otros factores.

Interacción:

Interacción es el fracaso de los niveles de un solo factor de responder de la misma forma frente a los niveles de otro factor o factores. Cuando dos factores interactúan la respuesta a los cambios de un factor está condicionado por el nivel de otro factor.

Factorial simétrico:

Es aquel en donde cada factor presenta igual número de niveles, pudiéndose presentar mediante la notación p^n, donde ***p*** se refiere al número de niveles y ***n*** al número de factores.

Factorial asimétrico:

Es aquel en donde cada factor presenta desigual número de niveles, pudiendo representarse como ***p*** *x* ***q***, donde ***p*** identifica los niveles del primer factor y ***q*** los niveles del segundo factor.

5.2 Ventajas

a. Se logra una gran eficiencia en el uso de los recursos experimentales disponibles.

b. Se obtiene información respecto a las diversas interacciones.

c. Los resultados experimentales son aplicables a un rango de condiciones más amplio debido a las combinaciones de los diversos factores en un solo experimento. Los resultados son de naturaleza más comprensiva.

d. Los experimentos factoriales son más eficientes que los experimentos simples.

5.3 Desventajas

a. El resultado del experimento y el análisis estadístico resultante son más complejos.

b. Con un gran número de combinaciones de tratamientos, la relación de unidades experimentales homogéneas es más difícil.

c. Convencidos de que algunas de las combinaciones de tratamientos pueden ser de muy poco o ningún interés, algunos de los recursos experimentales puden ser malgastados.

d. El número de tratamientos o combinaciones aumentan rápidamente.

5.4 Cuando Usarlos

La elección de un experimento factorial responderá a la necesidad de evaluar simultáneamente dos o más factores. Ejemplos:

a. Variedades y dosis de fertilizantes (2 factores)

b. Distancias de siembra, niveles de nitrógeno (2 factores)

c. Dosis, épocas y formas de aplicación de diferentes hormonas (3 factores)

5.5 Modelo Estadístico

El modelo que se describe corresponde a un experimento bifactorial, en arreglo combinatorio dispuesto en un diseño en bloques completos al azar, debido a que es el más usado.

$$Y_{ijk} = \mu + A_i + B_j + AB_{ij} + \beta_k + \varepsilon_{ijk}$$

donde:

Y_{ijk} = variable de respuesta asociada a la ijk - ésima unidad experimental

μ = Efecto de la media general

A_i = Efecto del i - ésimo nivel del factor "A"
B_j = Efecto del j - ésimo nivel del factor "B"

AB_{ij} = interacción del i - ésimo nivel del factor "A" con el j - ésimo nivel del factor "B"

β_k = Efecto del k - ésimo bloque

ε_{ijk} = error experimental asociado a la ijk - ésima unidad experimental

5.6 Análisis de Varianza

5.6.1 Hipótesis

Ho: $A_i = 0$ contra,

Ha: No todos los $A_i = 0$

Ho: $B_j = 0$ contra,

Ha: No todos los $B_j = 0$

Ho: $(AB)_{ij} = 0$ contra,

Ha: No todos los $(AB)_{ij} = 0$

Con mucha frecuencia éstas son las hipotesis que se deben probar.

5.6.2 Tabla de Datos

(Factorial 3 * 2 en bloques al azar)

Bloques	A_1		A_2		A_3		
	B_1	B_2	B_1	B_2	B_1	B_2	$Y_{..k}$
I	Y_{111}	Y_{112}	Y_{121}	Y_{122}	Y_{131}	Y_{132}	$Y_{..1}$
II	Y_{211}	Y_{212}	Y_{221}	Y_{222}	Y_{231}	Y_{232}	$Y_{..2}$
III	.	.	.	.	.	.	.

.	.	.	.	.	.	.	.
r	Y_{r11}	Y_{r12}	Y_{21r}	Y_{22r}	Y_{31r}	Y_{32r}	$Y_{..r}$
$Y_{ij.}$	$Y_{11.}$	$Y_{12.}$	$Y_{21.}$	$Y_{22.}$	$Y_{31.}$	$Y_{32.}$	$Y_{...}$

5.6.3 Tabla adicional para factores

Factor "B"	Factor "A"			$Y_{.j.}$
	A_1	A_2	A_3	
B_1	$Y_{11.}$	$Y_{21.}$	$Y_{31.}$	$Y_{.1.}$
B_2	$Y_{12.}$	$Y_{22.}$	$Y_{32.}$	$Y_{.2.}$
$Y_{i..}$	$Y_{1..}$	$Y_{2..}$	$Y_{3..}$	$Y_{...}$

$Y_{..k}$ = Total de bloques
$Y_{ij.}$ = Total de tratamientos
$Y_{i..}$ = Total del factor "A"
$Y_{.j.}$ = Total del factor "B"
$Y_{...}$ = Gran total

5.6.4 Tabla de Análisis de Varianza

FV	gl	SC	CM	FC
Bloques	r – 1	$\sum_{k=1}^{r}\frac{Y_{..k}^2}{ab}-\frac{Y_{...}^2}{abr}$		
Tratamientos	ab – 1	$\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{Y_{ij.}^2}{r}-\frac{Y_{...}^2}{abr}$	SC_{trat}/gl_{trat}	CM_{trat}/CM_{ee}

A	a – 1		SC_A/gl_A	CM_A/CM_{ee}
B	b – 1		SC_B/gl_B	CM_B/CM_{ee}
AB	(a - 1) (b - 1)	$SC_{TOTAL}\text{-}SC_A\text{-}SC_B$	SC_{AB}/gl_{AB}	CM_{AB}/CM_{ee}
Error	ab - 1 (r - 1)	$SC_{TOTAL}\text{-}SC_{BLOQUES}\text{-}SC_{TRAT}$	Sc_{ee}/gl_{ee}	
Total	abr – 1	$\sum_{i=1}^{a}\frac{Y_{i..}^2}{rb}-\frac{Y_{...}^2}{abr}$		

$$\sum_{j=1}^{b}\frac{Y_{.j.}^2}{ra}-\frac{Y_{...}^2}{abr}$$

Los valores de Fc que aparecen en la tabla sirven para probar hipótesis de la siguiente manera:

1. *Para factor "A"*

   $Fc = CM_A/CM_{ee}$; $F_{tab\,(a-1;\ glee\ ;\ \alpha)}$

2. *Para factor "B"* $\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}Y_{ijk}^2-\frac{Y_{...}^2}{abr}$

   $Fc = CMB_B/CM_{ee}$; $F_{tab\,(b-1;\ glee\ ;\ \alpha)}$

3. *Para la interacción "AB"*

   $Fc = CM_{AB}/CM_{ee}$; $F_{tab\,(a-1,\ b-1\ ;\ glee\ ;\ \alpha)}$

5.6.5 Regla de Decisión

Si $Fc \geq Ft$ Se rechaza la hipótesis nula
Si $Fc < Ft$ No se rechaza la hipótesis nula

Para efectuar la prueba de Tukey o SNK seguir el procedimiento anteriormente descrito, considerando los comparadores de la manera siguiente:

$$Wp = q_{(b,glee,a)} * S\bar{x} \qquad S\bar{x} = \sqrt{\frac{CMee}{ra}}$$

1. Para factor "A"

donde: $$Wp = q_{(ab,glee,a)} * S\bar{x} \qquad S\bar{x} = \sqrt{\frac{CMee}{r}}$$

q = valor tabular (Consultar tabla de Tukey)

a = niveles del factor A.

$$Wp = q_{(a,glee,a)} * S\bar{x} \qquad S\bar{x} = \sqrt{\frac{CMee}{rb}}$$

α = nivel de significancia

S_x = error estandar

CMee= cuadrado medio del error

r = número de repeticiones

b = número de niveles del factor B.

2. Para factor "B"

3. Para la interacción AB

A quienes hacerle la prueba de Tukey: ver tabla que aparece en el siguiente tema de parcelas divididas (Sección 6.7).

5.7 Ejemplo de Aplicación

Se estudió el efecto de diferentes dosis de fertilizante fosforado sobre dos tipos de planta de haba *Vicia faba*. Se pensó que los tipos de planta bien podían responder en forma diferente a la fertilización, así que se decidió llevar a cabo un experimento factorial con dos factores:

Factor	Nivel		
	1	2	3
A. Tipo de Planta	Corto	Alto	
B. Dósis de Fósforo	0 Kg/ha P_2O_5	25 Kg/ha P_2O_5	50 Kg/ha P_2O_5

El experimento se llevo a cabo mediante un diseño en bloques completos al azar en arreglo combinatorio, con cuatro repeticiones. Los resultados (producción en kg/ha) se aprecian en el cuadro 4.

Cuadro 4: Producción (kg/ha) de dos tipos de haba bajo diferentes niveles de fertilización de fósforo.

Tipo de Planta	Dosis de Fósforo	Bloques			
		I	II	III	IV
Corto	0	11.5	13.6	14.3	14.5
	25	17.1	17.6	17.6	18.1
	50	18.2	17.6	18.2	18.9
Largo	0	11.0	11.2	12.1	12.6
	25	8.3	10.5	9.1	12.8

50	15.7	16.7	16.6	17.5

5.7.1 Programa de SAS

A continuación el programa SAS y la salida que considera el análisis de varianza con los dos factores analizados.

```
OPTIONS NODATE;
DATA dos;
INPUT bloque planta $ fosforo rend;
CARDS;
1 Corto  0   11.5
1 Corto 25  17.1
.
.
4 Largo 25  12.8
4 Largo 50  17.5
;
PROC GLM;
    CLASS bloque planta fosforo;
    MODEL rend=bloque planta fosforo planta*fosforo;
    LSMEANS planta*fosforo/PDIFF;
RUN;
```

5.7.2 Salida de SAS

```
                               The SAS System
                        General Linear Models Procedure
                            Class Level Information

                         Class     Levels    Values

                         BLOQUE         4    1 2 3 4

                         PLANTA         2    Corto Largo

                         FOSFORO        3    0 25 50

                     Number of observations in data set = 24


                                 The SAS System
                        General Linear Models Procedure

Dependent Variable: REND
                                    Sum of            Mean
Source                  DF          Squares          Square   F Value     Pr > F
```

```
Model                   8     234.70000000     29.33750000     50.72     0.0001

Error                  15       8.67625000      0.57841667

Corrected Total        23     243.37625000

                  R-Square             C.V.        Root MSE           REND Mean

                  0.964350         5.195813       0.7605371           14.637500


Source                 DF        Type I SS     Mean Square   F Value     Pr > F

BLOQUE                  3      13.32125000      4.44041667      7.68     0.0024
PLANTA                  1      77.40041667     77.40041667    133.81     0.0001
FOSFORO                 2      99.87250000     49.93625000     86.33     0.0001
PLANTA*FOSFORO          2      44.10583333     22.05291667     38.13     0.0001

Source                 DF      Type III SS     Mean Square   F Value     Pr > F

BLOQUE                  3      13.32125000      4.44041667      7.68     0.0024
PLANTA                  1      77.40041667     77.40041667    133.81     0.0001
FOSFORO                 2      99.87250000     49.93625000     86.33     0.0001
PLANTA*FOSFORO          2      44.10583333     22.05291667     38.13     0.0001
```

```
                                 The SAS System

                        General Linear Models Procedure
                              Least Squares Means

                     PLANTA   FOSFORO         REND   LSMEAN
                                            LSMEAN   Number
                     Corto    0         13.4750000     1
                     Corto    25        17.6000000     2
                     Corto    50        18.2250000     3
                     Largo    0         11.7250000     4
                     Largo    25        10.1750000     5
                     Largo    50        16.6250000     6

                        Pr > |T| H0: LSMEAN(i)=LSMEAN(j)

                 i/j     1       2       3       4       5       6
                 1    .      0.0001  0.0001  0.0053  0.0001  0.0001
                 2  0.0001    .      0.2633  0.0001  0.0001  0.0899
                 3  0.0001  0.2633    .      0.0001  0.0001  0.0094
                 4  0.0053  0.0001  0.0001    .      0.0114  0.0001
                 5  0.0001  0.0001  0.0001  0.0114    .      0.0001
                 6  0.0001  0.0899  0.0094  0.0001  0.0001    .
```

6. ARREGLO EN PARCELAS DIVIDIDAS

6.1 Características del Arreglo

Frecuentemente por razones de nuestro interés recurrimos a un experimento factorial que pudiera representar dificultades de manejo en algunas ocasiones; por ejemplo si se quiere evaluar:

FACTOR	NIVELES		
A. Sistemas de Riego	Goteo	Aspersión	Gravedad
B. Niveles de nitrógeno	0	80	Kg/ha

Cuando se hace la distribución en el campo surge el inconveniente que presenta la aplicación de los tratamientos a las unidades experimentales; lo que implica demasiado trabajo. Bajo estas circunstancias se justifica el uso del arreglo en parcelas divididas. En este ejemplo el factor sistemas de riego es el que ilustra este incoveniente.

A continuación se presenta la forma como quedarían los tratamientos al usar un arreglo combinatorio en bloques al azar.

BLOQUE I

Goteo	Aspersión	Gravedad	Goteo	Gravedad	Aspersión
0 Kg/ha	80 Kg/ha	0 Kg/ha	80 Kg/ha	80 Kg/ha	0 Kg/ha

Lo que se pretende al emplear un arreglo en parcelas divididas es que los niveles del factor problema se asignen a *parcelas grandes* (principales), éstos pueden disponerse en cualquiera de los diseños ya estudiados (bloques al azar, cuadrado latino ,etc).

Los niveles del segundo factor se asignan aleatoriamente a las *parcelas pequeñas* (sub-parcelas) dentro de cada parcela grande.

En conclusión se recomienda el uso de parcelas divididas cuando nos represente un ahorro de trabajo grande o haga más práctica la conducción del experimento.

Otras de las razones más comunes para usar parcelas divididas son:

a. Permite un uso eficiente de algunos tratamientos que necesitan por su naturaleza, tamaños de parcela mayores que otros factores; tal es el caso de factores como sistemas de riego, manejo de suelos, temperaturas de cuartos fríos, aspersiones, sistemas de siembra.

b. Proporciona una mayor precisión en la estimación de los efectos de algunos factores comparados con otros. Por ejemplo, si los patrones de un cultivo de cítricos se ubica en las parcelas principales y las variedades en las subparcelas, el efecto de variedades se estima con mayor precisión.

c. Permiten la introducción de nuevos tratamientos dentro de un experimento que ya está en ejecución.

BLOQUE I usando parcelas divididas

Gravedad	Goteo	Aspersión
0 Kg/ha	80 Kg/ha	0 Kg/ha
80 Kg/ha	0 Kg/ha	80 Kg/ha

Parcela Grande   Parcela Pequeña

6.2 Ventaja

a. Facilita el manejo del experimento

6.3 Desventajas

a. Análisis estadístico más elaborado.

b. Se pierde precisión en la estimación de los efectos promedio de los tratamientos asignados a las parcelas grandes; suele frecuentemente aumentar la precisión en la estimación de los efectos de los tratamientos asignados a las parcelas pequeñas.

La razón fundamental de lo anterior se debe a que el error para parcela grande suele ser mayor que el error para parcela pequeña. Se sugiere entonces colocar el factor más importante en parcela pequeña y el otro en la parcela grande.

6.4 Aleatorización

La distribución de los tratamientos a las parcelas grandes (principales) se hace de la misma forma que se sugiere para el correspondiente diseño que se use (bloques al azar, cuadrado latino, u otro).

Los tratamientos de las parcela pequeñas (sub-parcelas) se asignan aleatoriamente dentro de cada parcela principal en forma independiente para cada una de éstas últimas.

6.5 Modelo Estadístico (Para un diseño en bloques al azar)

$$Y_{ijk} = \mu + \beta_j + \alpha_i + \xi_{ij} + \rho_k + \alpha\rho_{ik} + \varepsilon_{ijk}$$

donde:

Y_{ijk} = Variable de respuesta de la ijk - ésima unidad experimental

μ = Efecto de la media general

β_j = Efecto del j - ésimo bloque

αi = Efecto del i - ésimo nivel del factor A.

ξ_{ij} = Error experimental asociado a la parcela grande

ρ_k = Efecto del k - ésimo nivel del factor B

$\alpha\rho_{ik}$= Efecto debido a la interacción del j - ésimo nivel del factor A con los k - ésimos niveles del factor B.

ε_{ijk} = Error experimental asociado a la parcela pequeña.

6.6. Análisis de Varianza

6.6.1 Hipótesis

Ho: $A_i = 0$ contra,

Ha: No todos los $A_i = 0$

Ho: $B_j = 0$ contra,

Ha: No todos los $B_j = 0$

Ho: $(AB)_{ij} = 0$ contra,

Ha: No todos los $(AB)_{ij} = 0$

Con mucha frecuencia éstas son las hipotesis que se deben probar.

6.6.2 Tabla de Datos

Parcela Grande	Parcela pequeña	BLOQUES I	II	III	$Y_{.jk}$
A_1	B_1	Y_{111}	Y_{121}	Y_{311}	$Y_{.11}$
	B_2	Y_{112}	Y_{122}	Y_{312}	$Y_{.12}$
	B_3	Y_{113}	Y_{123}	Y_{313}	$Y_{.13}$
Total parcela grande		$Y_{11.}$	$Y_{21.}$	$Y_{31.}$	$Y_{1..}$
A_2	B_1	Y_{211}	Y_{221}	Y_{231}	$Y_{.21}$
	B_2	Y_{212}	Y_{222}	Y_{232}	$Y_{.22}$
	B_3	Y_{213}	Y_{223}	Y_{233}	$Y_{.23}$
Total parcela grande		$Y_{21.}$	$Y_{22.}$	$Y_{23.}$	$Y_{2..}$
Total bloques		$Y_{.1.}$	$Y_{.2.}$	$Y_{.3.}$	$Y_{...}$

Tabla adicional: Esta taba es útil para calcular la suma de cuadrados para ambos factores.

Factor "B"	Factor "A" A_1	A_2	Total
B_1	$Y_{1.1}$	$Y_{2.1}$	$Y_{...1}$
B_2	$Y_{1.2}$	$Y_{2.2}$	$Y_{...2}$
B_3	$Y_{1.3}$	$Y_{2.3}$	$Y_{..3}$
Total	$Y_{1..}$	$Y_{2..}$	$Y_{...}$

6.6.3 Tabla para el Análisis de Varianza (Bloques al azar)

Fuente de variación	Grados de libertad	Suma de Cuadrados	C. medios	Valor de F calc.

Bloques	r – 1			
A	a – 1		SC_A/GL_A	$CM_A/CM_{(a)}$
Error (a)	(a - 1) (r - 1)	$SC_{subt} - SC_{bloque} - SC_A$	$SC_{(a)}/GL_{(a)}$	
Subtotal	ra – 1			
B	b – 1		SC_B/GL_B	$CM_B/GL_{(b)}$
		$\sum_{j=1}^{r} \frac{Y_{.j.}^2}{ab} - \frac{Y_{...}^2}{arb}$		
AB	(a - 1) (b - 1)		SC_{AB}/GL_{AB}	$CM_{AB}/CM_{(b)}$
		$\sum_{i=1}^{a} \frac{Y_{i..}^2}{rb} - \frac{Y_{...}^2}{arb}$		
Error (b)	a (b - 1) (r -1)	$SC_{tot} - Sc_{subt} - SC_B - SC_{AB}$	$SC_{(b)}/GL_{(b)}$	
Total	abr – 1	$\sum_{j=1}^{r} \sum_{i=1}^{a} \frac{Y_{ij.}^2}{b} - \frac{Y_{...}^2}{arb}$		
		$\sum_{k=1}^{b} \frac{Y_{..k}^2}{ar} - \frac{Y_{...}^2}{arb}$		

6.7 Prueba de Tukey o SNK

En la siguiente tabla se indican a qué fuente de variación hay que efectuarle dicha prueba en relación a su significancia.

$$\sum_{j=1}^{r} \sum_{k=1}^{b} \frac{Y_{i.k}^2}{r} - \frac{Y_{...}^2}{arb} - SC_A - SC_B$$

Fuente de Variación	Resultado de la Prueba							
A	*	NS	*	NS	NS	*	*	NS
B	*	*	NS	NS	*	NS	*	NS
AB	*	*	*	*	NS	NS	NS	NS

$$\sum_{j=1}^{r} \sum_{i=1}^{a} \sum_{k=1}^{b} Y_{ijk}^2 - \frac{Y_{...}^2}{arb}$$

A quién hacerle prueba	AB	AB	AB	AB	B	A	A ; B	----

$$W = t_{ab} * S\bar{x}$$

***** : Significativo
NS : No significativo

$$t_{ab} = \frac{(b-1)\, CM(b)\, t_b + CM(a)\, t_a}{(b-1)\, CM(b) + CM(a)} \qquad S\bar{x} = \sqrt{\frac{(b-1)\, CM(b) + CM(a)}{rb}}$$

6.7.1 Prueba de Tukey para el factor de la Parcela Grande

6.7.2 Prueba de Tukey para el factor de la parcela pequeña

$$W = q_{(a,gl(a),a)} * S\bar{x} \qquad S\bar{x} = \sqrt{\frac{CM(a)}{rb}}$$

6.7.3 Prueba de Tukey para la interacción "AB"

6.7.3.1 Comparación de tratamientos dentro de una ***misma*** *parcela grande.*

$$W = q_{(b,gl(b),a)} * S\bar{x} \qquad S\bar{x} = \sqrt{\frac{CM(b)}{rb}}$$

6.7.3.2 Comparación de tratamientos entre parcelas grandes ***diferentes****.*

, donde,

$$W = q_{(ab,gl(b),a)} * S\bar{x} \qquad S\bar{x} = \sqrt{\frac{CM(b)}{rb}}$$

$t_a = gl(a), a$ $\qquad$ $t_b = gl(b), a$ *(Valores de tabla de "t" de Student –2 colas-)*

6.8 Ejemplo de Aplicación

En un experimento se probaron dos métodos de siembra y tres variedades de arroz. Por razones de facilidad en el manejo del mismo, se decidió utilizar un arreglo en parcelas divididas, en donde las parcelas grandes correspondieron a los métodos de siembra y en las parcelas pequeñas se sembraron las tres variedades de arroz. Los tratamientos se distribuyeron mediante un diseño en bloques completos al azar con 4 repeticiones. La variable de respuesta fué kg/parcela de 72 m^2. A continuación se detallan los factores evaluados.

Métodos de siembra:	S: Secano	I: Inundación	
Variedades:	BB: Blue Bonnett	BP: Belle Patna	CR: Criolla

Cuadro 5: Rendimiento de arroz en kg/parcela de 72 m^2 , al comparar distintos métodos de siembra en diferentes variedades de arroz.

Métodos de siembra	Variedades	Bloques			
		I	II	III	IV
S	BB	8.5	8.2	8.0	7.6
	BP	8.0	7.0	7.3	6.9
	CR	10.0	10.8	11.0	11.6
I	BB	5.8	5.6	5.2	5.8
	BP	7.0	7.1	7.3	7.3
	CR	8.3	8.0	8.0	8.7

6.8.1 Programa de SAS

A continuación se presenta el programa SAS y la salida que considera el análisis de varianza con arreglo en parcelas divididas.

```
OPTIONS NODATE;
DATA dos;
INPUT bloque metodo $ variedad $ rend;
CARDS;
1 S BB 8.5
1 S BP 8.0
1 S CR 10.0
```

```
1 I BB 5.8
1 I BP 7.0
1 I CR 8.3
2 S BB 8.2
.
.
4 S CR 11.6
4 I BB 5.8
4 I BP 7.3
4 I CR 8.7
;
PROC GLM;
   CLASS bloque metodo variedad;
   MODEL rend=bloque metodo variedad metodo*variedad bloque*metodo;
   TEST H=metodo E=bloque*metodo;
   LSMEANS metodo*variedad/PDIFF E=bloque*metodo;
RUN;
```

6.8.2 Salida de SAS

```
                                 SAS

                    General Linear Models Procedure
                        Class Level Information

                      Class    Levels    Values

                      BLOQUE        4    1 2 3 4

                      METODO        2    I S

                      VARIEDAD      3    BB BP CR

                Number of observations in data set = 24


                                SAS
                    General Linear Models Procedure

Dependent Variable: REND
                                       Sum of           Mean
Source                  DF          Squares         Square   F Value     Pr > F

Model                   11      60.48833333     5.49893939     24.11     0.0001

Error                   12       2.73666667     0.22805556

Corrected Total         23      63.22500000

                  R-Square             C.V.       Root MSE           REND Mean

                  0.956715         6.064148       0.477552          7.87500000


                                 SAS

                    General Linear Models Procedure

Dependent Variable: REND

Source                  DF        Type I SS    Mean Square   F Value     Pr > F

BLOQUE                   3       0.17500000     0.05833333      0.26     0.8558
METODO                   1      18.02666667    18.02666667     79.05     0.0001
VARIEDAD                 2      34.30750000    17.15375000     75.22     0.0001
METODO*VARIEDAD          2       7.77583333     3.88791667     17.05     0.0003
BLOQUE*METODO            3       0.20333333     0.06777778      0.30     0.8268

Source                  DF      Type III SS    Mean Square   F Value     Pr > F

BLOQUE                   3       0.17500000     0.05833333      0.26     0.8558
METODO                   1      18.02666667    18.02666667     79.05     0.0001
VARIEDAD                 2      34.30750000    17.15375000     75.22     0.0001
METODO*VARIEDAD          2       7.77583333     3.88791667     17.05     0.0003
```

```
                                  SAS

                     General Linear Models Procedure

Dependent Variable: REND

Source                  DF     Type III SS    Mean Square   F Value     Pr > F

BLOQUE*METODO            3      0.20333333     0.06777778      0.30     0.8268

Tests of Hypotheses using the Type III MS for BLOQUE*METODO as an error term

Source                  DF     Type III SS    Mean Square   F Value     Pr > F

METODO                   1     18.02666667    18.02666667    265.97     0.0005


                                  SAS
                     General Linear Models Procedure
                           Least Squares Means

   Standard Errors and Probabilities calculated using the Type III MS for
                      BLOQUE*METODO as an Error term

                   METODO   VARIEDAD          REND   LSMEAN
                                            LSMEAN   Number

                   I        BB           5.6000000     1
                   I        BP           7.1750000     2
                   I        CR           8.2500000     3
                   S        BB           8.0750000     4
                   S        BP           7.3000000     5
                   S        CR          10.8500000     6


                                   SAS
                     General Linear Models Procedure
                           Least Squares Means

                Least Squares Means for effect METODO*VARIEDAD
                       Pr > |T| H0: LSMEAN(i)=LSMEAN(j)

Dependent Variable: REND

                i/j      1       2       3       4       5       6
                1    .      0.0034  0.0007  0.0009  0.0027  0.0001
                2  0.0034    .      0.0100  0.0164  0.5458  0.0003
                3  0.0007  0.0100    .      0.4119  0.0141  0.0008
                4  0.0009  0.0164  0.4119    .      0.0245  0.0006
                5  0.0027  0.5458  0.0141  0.0245    .      0.0003
                6  0.0001  0.0003  0.0008  0.0006  0.0003    .

NOTE: To ensure overall protection level, only probabilities associated with
      pre-planned comparisons should be used.
```